# PATENT ABSTRACTS OF JAPAN

(11)Publication number :

(43) Date of publication of application: 24.06.1992

04-177461

(51)Int.CI. G06F 15/31 G06F 15/332

(21)Application number: 02-305035 (71)Applicant: MITSUBISHI ELECTRIC CORP

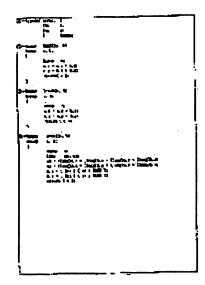
(22)Date of filing: 07.11.1990 (72)Inventor: KEGASA MITSUYOSHI

## (54) HIGH SPEED TECHNICAL CALCULATION METHOD

## (57)Abstract:

PURPOSE: To perform a technical calculation at a high speed without using any expensive numerical arithmetic processor and at the same time to reduce the necessary capacity of a memory byperforming a numerical operation through an integer arithmetic and in a block floating point format.

CONSTITUTION: A fixed floating point arithmetic using an integer arithmetic is expressed in a block floating point number that showed the data in an exponent part or a coefficient part common to all data and a fixed decimal point of each data. Then a block floating point arithmetic applying an integer arithmetic is carried out. It is supposed that a decimal, point is set at the 13th bit of an integer and then that 8192=213 of an integer is equivalent to 1.01 of a fixed decimal point. In this fixed decimal point format, the value covering -4.0 (minimum) through about 3.9999 (maximum) can be expressed. The expressible number most approximate to 0 is about 0.000122 and a number approximate to 1.0 has the decimal accuracy equivalent to about 4 digits. The diagram shows a program used for calculation of a complex number end applying a fixed decimal point in a C language.



## **LEGAL STATUS**

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

⑪特許出願公開

# ⑩ 公 開 特 許 公 報 (A) 平4-177461

Sint. Cl. ⁵

識別記号

庁内整理番号

43公開 平成4年(1992)6月24日

G 06 F 15/31 15/332 Z A 6798-5L 6798-5L

審査請求 未請求 請求項の数 1 (全10頁)

会発明の名称
高速技術計算方法

**负特 願 平2-305035** 

②出 願 平2(1990)11月7日

@発明者 毛笠 光容

兵庫県尼崎市塚口本町8丁目1番1号 三菱電機株式会社

通信機制作所内

⑪出 願 人 三菱電機株式会社

東京都千代田区丸の内2丁目2番3号

@代理人 弁理士 大岩 增雄 外2名

#### 明期中期

1. 発明の名称

高速技術計算方法

2. 特許請求の範囲

構造化された汎用の高級言語上で、複数個のデータを処理する技術計算において、整数演算を利用したプロック化浮動小数点演算もしくは固定小数点演算を用いたことを特徴とする、高速技術計算方法。

3. 発明の詳細な説明

[産業上の利用分野]

この発明は、コンピュータによる技術計算の高 速化に関するものである。

〔従来の技術〕

第5図(a), (b)は、CQ出版の「Cによる科学技術計算」小池慎一著の209ページ〜211ページに掲載されていたC首語による複素数演算プログラムである。

(1)は複素数を表現する形式complexを定義 しており、複素数の実数部を倍精度浮動小数点 (以降doubleと略す)のreであらわし、虚部をdoubleのiaであらわしている。(2)は(1)で 定義したcomplex型の加算用の関数caddであり、

(3) はcomplex型の減算用の関数csul、(4) はcomplex型の乗算用の関数csulであり、(5)

はcomplex型の除算用の関数cdivである。(6)

はcomplex型の数の絶対値を求める関数cabsであり、 (7) はdouble型の数2つを各成分とするcomplex型の数を求める関数to-complexである。

第6図は同じく「Cによる科学技術計算」の21 8ページに掲載されていた複素数複算を用いたFFT プログラムである。(10)はサブルーチン名 cfftの宣言部、(11)はピットリバーサル操作 部の先頭、(12)はバタフライ積部の先頭を示 している。

第7図(a)、(b)は第6図を元にして図6のプログラムで省略されていたFFT演算後の割り算部を付加した実用的な複類数FFTプログラムである。(10)~(12)は第6図と同じもので

あり、 (13) はFFT演算後の配列 Z の要素を n で割る割り算部を示している。 第8図は、ビットリバーサル部のフローチャート、第9図はバタフライ積部のフローチャートである。

次に動作について説明する。

第5図の(2)の関数caddは複素数の変数 a,bを引き数として受け取り、その和である複素数 2を返すもので、

Z.re(Zの実部)= a.re(aの実部)+b.re(bの実部)
Z.im(Zの虚部) = a.im(aの虚部)+b.im(bの虚部)
を計算している。

第5図の(3)の関数csubは同様に複楽数の差を返すもので、X.re=a.re-b.re, X.im=a.im-b.imを計算している。

第5図の(4)の関数caulは同様に複素数の積 を返すもので、

X.re=a.re×b.re-a.lm×b.lm X.im=a.re×b.im-a.lm×b.re を計算している。

第9図は前記プログラムのバタフライ演算部のフローチャートであり、第7図の(12)の部分を説明している。

第10図は第7図の(13)(第6図では省略 されている)の割り算部のフローチャートであ る

第7図のFFTプログラム cfftに複素数データの配列 2、FFTの次数を表わす整数 1、FFTと RFTを選択する整数の変数 f (f=1 なら RFT、f=-1 なら FFTが実行される)を引数として与えて実行させると、f=1 なら配列 Z のデータをフーリエ逆変換したデータが配列 Z に求められる。 f=-1 なら、配列 Z のデータをフーリエ変換したデータが配列 Z に求められる。

FFTプログラムは、離数フーリエ変換を高速で 行なうプログラムで時間的に変化する離数データ g (k)の周波数スペクトルG (n)を、

$$G(n) = \frac{1}{N} \sum_{k=0}^{N-1} g(k) W_{k}^{nk}$$

という関係を用いて求めるプログラムである。但

第5図の(5)の関数cdivは同様に複素数の商を返すもので、

$$\chi. re = \frac{a. re \times b. re + a. im \times b. im}{b. re \times b. re + b. im \times b. im}$$

を計算している。又、分母がOの時はエラー表示をするようになっている。

第5図の (6) の関数 cabs は a の絶対値をdouble で返す関数で | a | =√(a.re) <sup>2</sup>+ (a.is) <sup>1</sup>を 計算している。

第 5 図の(7)の関数 to complex は doubleの変数 x と y を 引き数として受け取り、これを 複素数 2 に変換して返す関数で Z.re=x, Z.im=y を 実行している。

第6図、第7図は複素数演算を用いたFFTのプログラムの例である。

第8図は前配のプログラムのビットリパーサル 部のフローチャートであり、第7図の(11)の 部分を説明している。

し、Nはデータ数で通常 Z のべキ乗で表わされる数 (2,4,8,・・・,128,256,512,1024,・・・) とし、W w は e <sup>-1</sup> 持で表わされる数である。

FFTプログラムは主に電気倡号や機械的倡号などがどのような周波数成分を持つかを求めるのに使われており、古くは大型コンピュータでしか使われていなかったが、近年ではパソコン上や測定器上でも使われるようになっている。

第5図の(2)~(7)の複素数演算を使用して複素数FFTを記述すると、第7図のように短いプログラムとなる。

第7図のプログラムをパソコンPC9801RX2上でポーランド社のCコンパイラTurbo-Cを用いて実行させると、データ数N=1024の時、数値演算プロセッサがない場合で約13秒必要であった。

数値演算プロセッサ180287-10を付けると、同 じ演算が約3秒で実行できる。但し180287-10は 現状では4万円から6万円程度と高価である。

[発明が解決しようとする課題]

従来の技術計算方法は、以上のようにC言語で

この発明は、上配のような問題点を解消するためになされたもので、高価な数値演算プロセッサを使わずに、短時間で結果を求めることができ、かつ必要なメモリの容量も少ない高速技術計算プログラムを得ることを目的とする。

#### (無顧を解決するための手段)

この発明に係る高速技術計算方法は、データを 固定小数点またはブロック浮動小数点で表わし、

型 (以降intと略す) を概準としている。 1 6 ビットの符号付き整数では 2 <sup>1 4</sup>通りの数が表現でき 最小-32768~最大32767が表現できる。

本実施例ではこの整数の13ビット目に小数点があるものと仮定して整数の8192=2<sup>13</sup>が固定小数点の1.01に相当するものと設定した。この固定小数点フォーマットでは、最小-4.0~最大約3.8889が表現できる。表現できる0に最も近い数は約0.000122であり1.0付近の数なら10進で約4桁分の精度がある。

この固定小数点での演算は、加減算なら整数と全く同じ演算で実行できるが、乗算では1.0×1.0 =1.0となる様に調整する必要がある。又、16ビットの整数の積は、最大約10億となり16ビット符号付整数では表現できなくなる。そこで乗算では、各整数を32ビット符号付整数では最小約一2億~最大約21億までの数が表現できる。long intで乗算後8192で割り、int型に戻してやれば乗算ができ

整数演算を利用した固定小数点演算もしくは、ブロック浮動小数点で計算を行なうものである。

#### (作用)

この発明における高速技術計算方法は、必要以上に高精度であり多くのメモリを必要としし、演算に長い時間を要する倍精度浮動小数点演算を利用した固定小数点演算と各手の指数的又は係数的とそで表わしたがロックの化力が、である。 小数点演算で行なうことで技術計算を高速をかかった。 小数点演算で行なった必要なメモリー容量を削っている。

#### 〔寒施例〕

以下この発明の一実施例を説明する.

本実施例では、固定小数点を整数を使って表現している。整数はコンピュータによって表現が異なるが、例えばパーソナルコンピュータPC9801シリーズ用のCコンパイラTurbo-C ver2.0 (ボーランド社)では、整数は16ビットの符号付き整数

る。例えば、固定小数点数 a = 2.0, b = 0.5の積 を固定小数点のCに代入する演算をC 督語で表記 すると、

int a,b,c;

a=(int)(2.0\*8192.0);

b=(int)(0.5\*8192.0);

 $c=(int)((long\ int)a*(long\ int)b/8192);$ 

となる。 但し(int)は intへの型変換を指示するキャスト演算子である。  $又(long\ int)$ は  $long\ int$ への型変換を指示するキャスト演算子である。 固定小数点の割り算も同様に  $long\ int$ で計算する必要がある。 例えば a=1.0, b=2.0で a 割る b を c に代入するには

int a,b,c;

a=(int)(1.0\*8192.0);b=(int)(2.0\*8192.0); c=(int)((long int)a\*8192/(long int)b); とC首語で記述できる。

このように固定小数点の演算を整数演算の組み 合わせで実行することが可能である。

さらに複素数を2つの固定小数点数で表現する

ことも可能である。つまり複素数の実部を1つの 固定小数点数で、虚部をもう1つの固定小数点数 で表現することができる。これをC質語でbcomp 型と定義すると以下のようになる。

typedef struct(int x; int y;)bcomp; bcomp Z1,Z2,Z3;

この場合複素数 Z 1 の実部が Z 1.xに、虚部が Z 1.yに対応する。 Z 1 = 0.7 + j 0.7, Z 2 = 0.5 + j 0.8 の和を Z 3 に代入するには上の定義式にて宣言した後

Z1. x=(int)(0.7\*8192.0);

Z1.y=(int)(0.7\*8192.0);

Z2. x=(int)(0.5\*8192.0);

Z2. y = (int) (0.8 \* 8192.0);

23. x=21. x+22. x;

Z3. y=Z1. y+Z2. y:

と番けば良い。ここで、わかるように浮動小数点型数から固定小数点数を利用した複素数への変換や、複素数の計算をそのまま記述すると大変見にくくかつ書くのも大変であるので、C言語の関数

割れば良いのであるが、分子は 2 つの固定小数点数の積の和であり、例えば (2.0+j0.0)/(2.0+j0.0) (0.0+j0.0)/(2.0+j0.0) (0.0+j0.0)/(2

除算をさらに高速化するには、分子への積をやめて分母を8192で割れば良いが、この場合は除数の絶対値が整数で√8192年90固定小数点で0.01より小さいとエラーとなる。

としてやることが好ましい。

第1図(a)、(b)はC言語で固定小数点を利用した複素数計算用のプログラムである。最初の宣言部は上記説明のものと全く同じである。 (2)の関数bcaddは加算用の関数であり、従来の実施例の関数caddのcomplexの部分をbcompに変えたものである。

- (3)の関数 bcsubは減算用の関数であり、これも従来の実施例の関数 csubの complexの部分をbcompに変えたものである。
- (4)の関数 beaulは積を求める関数であり、 従来の実施例の関数 caulの coaplexの部分を becap に変え、さらに実部、虚部各部の乗算を求める左 辺で整数の型を intから long intに変換してから 乗算を行ない固定小数点に戻す為、8192で割って から型を int型に戻すように書き変えている。
- (5)の関数 bcdivは商を求める関数であり、 従来の実施例の関数 cdivの complexを bcompに変え、整数の型変換を行なっている。固定小数点の 商であるので被除数へ8192を乗じてから、除数で

除算の特度を上げるには、たとえば被除数及び 除数の絶対値を1以下という条件にしてやり、 2<sup>31</sup>+(8192)<sup>2</sup>=32の余裕にして分子を32倍し、 分母を256で割ってから割り算を行なえば良い。 この場合、除数の絶対値が整数で√256=16,固定 小数点で0,002より小さい時エラーとなる。

第1図において(6)は、複素数の絶対値を求める関数bcabsである。従来の実施例のcabsのcom-plexの部分をbcompにおき変えたものである。

第2図において(7)は、絶対値が2以下のdoubleの数2つをそれぞれ実部,虚部とするbcompの数を求める関数tobcompであり、従来の実施例のtocomplexとほぼ同じであるが、固定小数点フォーマットに直す為に8192を掛けている。

- (8) は複素数 bcompの数の配列 Z ( ) の n 個の要素を 2 で割るサブルーチン bcnormであり、 後で述べる FFT プログラムで使用している。
- (9) は同じくbcompの配列 Z ( ) の n 個 1 要素をその各成分(実部、虚部)の絶対値の最大 が 1 以下になるように闘整するサブルーチン

bcnora2である。絶対値の最大値を1以下にする 為にス( )の全要素を変数 constで割り、配列 2 〔 )の共通の指数部に相当する変数 ampを con st倍している。このサブルーチンでは、乙と amp よりなるブロック浮動小数点フォーマットをサポ ートしている。

(10)は、上で述べた複素演算を利用した複素FFTプログラムである。 従来の実施例とほぼ同じアルゴリズムであるが、 従来の実施例ではデータ数 N での割り算をプログラムの最後でまとめて行なっていたが、 本プログラムではバタフライでは、 クリングラムでは、 アータを実行している。 FFTプログラムでは、 バタフライ積を1組行なう度にデータの最大値が約2倍に増える性質があるので固定小数点が途中でオーバーフローすることがないように毎回2で割ることにより、 N での割り算を実行させている。

この複素FFTプログラムのアルゴリズムの異なる部分を説明しているフローチャートを第11

プログラムでこのFFTプログラムで消費した時間 を求めさせている。

このメインプログラムを実行させて時間を測った結果1024点のFFTが約2秒で完了した。

従来の実施例に比べて約6倍の高速化が実現できた。数値演算プロセッサを使用したものよりも少し速い。但し精度は2桁~3桁程度である。例えば、8ビットのA/Dコンバータ等で入力されたデータの処理用としてはこの精度は充分ながであると思われる。又結果を顧面(400ドット×640ドット)にグラフィック表示するようなアプリケーションプログラムでも、これだけの精度があれば十分である。

なお、上記実施例ではC首語上でプログラムを 記述したがPASCAL等の他の構造化した高級言語で も関様のプログラムが作成できる。

又、上記実施例ではFFTを実行するプログラムを記述したが、同じような複数個のデータを処理する技術計算プログラムなら同様の手法で高速化が可能である。

図、第12図に示す。第11図において(13) は追加されたbcnoraサブルーチンの実行位置を示 している。第9図の従来のプログラムで®で示す Nでの割り算プログラムへ移行するかわりに、第 11図ではサブルーチンを終了してメインプログ ラムへ戻っている。

図12はbcnorsサブルーチンのアルゴリズムを 示すフローチャートで、配列2の各要素を2で割っている。この様に固点小数点フォーマットの複 素数bcmpを使って加減乗除算や、複素数FFTが実 現できる。

第4図は、上記の複素数FFTプログラムをテストする為に作られたC言語のメインプログラムであり、1024個のデータの実部がサインカーブを描くよう設定し、虚部は全てOになるよう設定し、複素数FFTプログラムを実行している。

つまり、時間軸で正弦振動をしているようなデータを入力した時の周波数スペクトラムを求めい ている。

pr\_timeというサブルーチンは時刻を表示する

又、上記実施例ではFFTを複素数演算を用いて 計算しているが、複素数の実数部と虚数部のデー タを分けてそれぞれ別の配列にして、通常の実数 演算を用いて計算することも可能である。

又、上記実施例で使用したデータの複素数表現、複素数演算を用いて各種の技術計算を行なうことも可能であるが、技術計算の種類によっては固定小数点の精度が10進数で約4桁である為、演算中に大きな誤差を生じる場合があるので注意が必要である。

### (発明の効果)

以上のようにこの発明によれば、数値演算を整数演算を利用してプロック浮動小数点フォーマットで行なったので、高価な数値演算プロセッサを使うことなく、高速で技術計算を行なうことができ、かつ必要とするメモリ容量も削減することができる。

### 4. 図面の簡単な説明

第1図はこの発明の一実施例による複素数の加減乗除と絶対値を求める関数の方法を示す図、第

## 特開平4-177461(6)

2 図はこの発明の一実施例による複素数の処理方 法を示す図、第3図はこの発明の一実施例による 後素 FFTを示す図、第4図はこの発明の一実施例 による図、第5図は従来の複素数演算用関数の 図、第6図は従来の複素数演算を用いたFFT図、 第7図は従来の実用的な複素数FFT図、第8図は 従来及びこの発明の一実施例共通の複素数FFTブ ログラムのビットリパーサル部のフローチャー ト、第9図は従来の複素数FFTプログラムのバタ フライ積部のフローチャート、第10図は従来の 祖素数FFTプログラムの割り算部のフローチャー ト、第11図はこの発明の一実施例による複楽数 FFTプログラムのバタフライ積部のフローチャー ト、第12図はこの発明の一実施例による複素数 FFT プログラムのサブルーチン BCNORMのフローチ ャートである。

(1)は複素数の構造体宜官、(2)は複素数の加算用関数、(3)は複素数の減算用関数、(4)は複素数の乗算用関数、(5)は複素数の 除算用関数、(6)は複素数の絶対値を求める関

```
数、 (7) は実数 2 つから複楽数を作る関数、 (8) は複楽数の配列の n 個の要素を 2 で割るサブルーチン、 (9) は複楽数の配列の要素の各成分の最大値を 1 より小さくするサブルーチン、 (10) は複素数 FFTサブルーチン、 (11) はピットリバーサル部、 (12) はパタフライ積部、 (13) は割り算部を示す。
```

なお、図中、同一符号は同一もしくは相当部分 を示す。

代理人 大岩增雄

```
D-bcomp bcd(v(a, b)
bcomp a. b;

{
    bcomp c:
    long x. xs. xy;
    x = (long)b.x = (long)b.x + (long)b.y = (long)b.y;
    x = (long)b.x = (long)b.x + (long)c.y = (long)b.y;
    xx = (long)a.y = (long)b.x + (long)c.x = (long)b.y;
    c.x = (int) ((xx / x) = 0);
    c.y = (int) ((xx / x) = 0);

}

return (c):

D-dcubla bcshs(a)
bcomp a:

{
    double x.y.x:
    x = (double)x.s:
    y = (double)x.y:
    x = sqri(x = x + y = y) / $192.0:
    return (x):

}

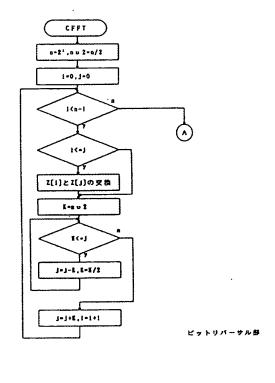
#### 1. CEU (b)
```

-400 -

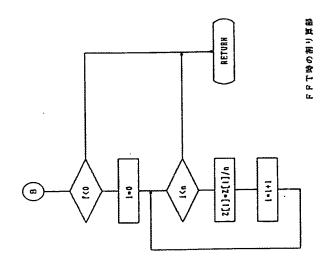
```
D-void
boxes
tat
                       effici.1.t)
                     P2;
i,f;
                     to: s.ev2.ip.i.j.k.e.le.let;
bccep i.e.u;
doubte c. e;
                     a * (int)pow(2,0,(double)1);
av2 * a / 2;
     -far (1=0,j=0; (Cs-1 ; t++)
                                 ( * x[1];
x[1] * x[1];
                     k --av2;
while ( k(-) )
      1
     -for ( s=1; s<=1 ; s++ )
                     le = (iel)pos(2,0,(double)s);
tei = te / 2;
u = tobcomy(1,0,0,0);
c = cos (s Pl / (double) lei );
s = ( (double) f ) = sis( N_Pl / (double) lei );
or = ( tobcomy(1,0,0);
for (j=0; j<1e/2; j++) {</pre>
                                    for ( !=!: ! < a : ! += !a )
                                             ip = [ + [al;
t = bcmul(x[tp],e);
x[ip] = bcmub(x[t],t);
x[i] + bcmdd(x[t],t);
                                   |
| = bcsel( u, u );
            O-bccore( 1, 1);
                                                             अ ३ छा
```

```
- complex cdiv(a,b)
              complex x:
              d-b, re + b, is=b, is:
if ( d==0 )
                     petacf("\7\7\7divided by zero error!!!\a");
d = TINT;
              z, re*(s, re*b, re*z, is*b, is)/d;
z, is*(s, is*b, re~s, re*b, is)/d;
  -double cabs(s)
complex s;
             reterm (sqrt(a, ress, ress, imes, im));
               if ( (a, re--0) &6 (a, tu--0) }
                            printf("\7\7\7 arg sero error!!!\n");
recere 10.0;
               else retura (atas2(a, in,s. re));
     complex comp(a)
                complex s;
                g.reme.re; s.temma.in;
reture %;
D-complex tocomiex(1,7)
double 1,7;
(
                complex 1:
2.re=1: 8.je=y:
reterm 1:
                                       # 5 EE (b)
```

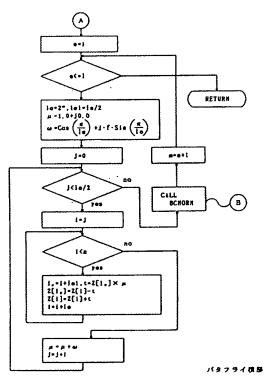
```
ist a, av2, ip, l, j, t, s, ie, lel; casples t, e, e; double s, c :
                n = ipow2(1): sv2=s/2 ;
 D-for (1-0,1-0: 1<a-l: 1++)
                                                      /* bit reversal operation */
                  it (164)
                             1-1[1]: 1[1]-1[1]: 1[1]-1:
                   k-sv2:
                   ebite ( k(-j )
                            1--h: k/-Z;
                   1++%:
      / ---- main loop *****/
  ⇔-for { e-1; e<-1; e++}
                  le- ipomE(a): iei=ie/E:
e = tocnapies(1.0, 0.0);
c = cos( f!//doable)iei );
e = teis[f/!/doable)iei);
e = tocnapies( c, e );
for (j=0; jcie/E: j**) ;
                               for (1-3: 1(n: 1+-1a)
                                           ip=|+ie1;
t =cmmi{z[ip],e): z[ip]=csub(z[i],l);
z[i]=cadd(z[i],l);
                               **casl( w. *):
                                           36 8 8
```

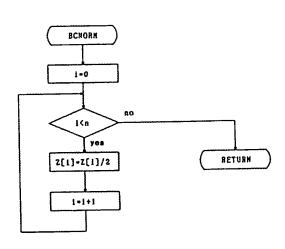


\* 8 B



河 0 区





第11回